
LONGITUDINAL EMPLOYER–HOUSEHOLD DYNAMICS

TECHNICAL PAPER NO. TP-2007-02

Social, Economic, Spatial, and Commuting Patterns of Informal Jobholders

Date: April 2007
Prepared by: Matthew R. Graham & Paul Ong
Contact: U.S. Census Bureau
4600 Silver Hill Rd.
LEHD Program
Attn: Jeremy S. Wu, Rm 6H141
Suitland, MD 20746 USA

This document reports the results of research and analysis undertaken by the U.S. Census Bureau staff. This document is released to inform interested parties of research and to encourage discussion. This research is a part of the U.S. Census Bureau's Longitudinal Employer-Household Dynamics Program (LEHD), which is partially supported by the National Science Foundation Grants SES-9978093 and SES-0427889 to Cornell University (Cornell Institute for Social and Economic Research), the National Institute on Aging Grant R01 AG018854, and the Alfred P. Sloan Foundation. The views expressed on statistical issues are those of the authors and not necessarily those of the U.S. Census Bureau, its program sponsors or data providers. Some or all of the data used in this paper are confidential data from the LEHD Program. The U.S. Census Bureau supports external researchers' use of these data through the Research Data Centers (see <<http://www.ces.census.gov/>>). For other questions regarding the data, please contact Jeremy S. Wu, Program Manager, U.S. Census Bureau, 4600 Silver Hill Rd., LEHD Program, Data Integration Division, Rm 6H141, Suitland, MD 20746, USA. (Jeremy.S.Wu@census.gov <<http://lehd.did.census.gov/>>).

Social, Economic, Spatial, and Commuting Patterns of Informal Jobholders^{*}
April 2007

Matthew R. Graham
Abt Associates and
U.S. Census Bureau

Paul Ong^{**}
School of Public Affairs
UCLA

* The authors wish to acknowledge the LEHD program and its staff for making this data available and providing support for these analyses. We would also like to thank Lucy Tran for her assistance in assembling the background materials, working on the graphics, and editing of this report.

** Corresponding Author: pmong@ucla.edu
School of Public Affairs, UCLA

Abstract

A significant number of employees within the United States can be considered “informal” or “off-the-books” workers. These workers, who by definition do not appear in administrative wage records, are distinct from the larger group of private jobholders who do appear in administrative records. However, while socioeconomic and spatial information on these individuals is readily available in standard datasets, such as the 2000 Decennial Census Long Form, it is not possible to identify the informal workers by only using such data because of the lack of accurate, formal wage records. This study takes advantage of firm-based data that originates in Unemployment Insurance administrative wage records linked with the Census Bureau’s household-based data in order to examine informal jobholders by their demographic characteristics as well as their economic, commuting, and spatial location outcomes. In addition this report evaluates whether informal jobholders should be included explicitly in future labor-workforce analyses and transportation modeling. The analyses in this report use the sample of workers who lived in Los Angeles County, California.

Keywords

Informal jobholders; off-the-book workers; informal economy; underground economy; labor-workforce; commuting patterns; modal split; transportation modeling; OD-matrix; Los Angeles; California; administrative wage records; US Census Bureau; Longitudinal Employer-Household Dynamics; LEHD.

Abbreviations

ACS	– American Community Survey
JH	– Jobholder
LEHD	– Longitudinal Employer-Household Dynamics
LF	– 2000 Decennial Census, Long Form Average 1-in-6 Sample
POR	– Place of Residence
POW	– Place of Work
PIK	– Protected Identification Key
PMSA	– Primary Metropolitan Statistical Area
SSN	– Social Security Number

INTRODUCTION

This is the second of three technical briefs examining workers who are difficult to study because of the unique nature of their labor market status: multiple jobholders, informal jobholders, and self-employed jobholders. This report covers the analysis of informal jobholders, who self-reported being privately employed but had no recorded wage earnings in employment data during the first two quarters of 2000. The analyses in these reports provide new insights by combining data from two sources: the 2000 Census and the LEHD Program, both of which are described later. Each dataset has its strengths (the Census data is population-based while the LEHD data is firm-based), and when combined, they provide complementing coverage and data items (Wu et al., 2005). The analyses have a number of objectives, but one of the most important is to examine the commute characteristics of informal workers relative to the commute characteristics for formal workers, who constitute the majority of the workforce. While there are some studies that examine the personal characteristics of these workers and their employment outcomes, there is a paucity of information on the commute itself: the travel mode, time, and distance, as well as origins and destinations. These characteristics are critical both to the study of the spatial structure of metropolitan labor markets and to transportation analysis and modeling.

The analysis in this report focuses on informal jobholders and compares them to the larger group of formal jobholders.* Studying informal workers is important because they contribute to the discrepancy in employment counts in LEHD data, which includes only those with paid jobs reported by firms in the unemployment insurance (UI) system. We can develop a partial profile of informal workers by merging LEHD data with the 2000 Decennial Census Long Form (LF) sample because this population-based survey includes those who were not included in the LEHD data. The coverage of this segment of the labor force in the LF sample is not known because some informal workers may be reluctant to report employment to the Census Bureau. Also adding error to the sample are informal jobholders who use others' social security numbers because of legal issues. Despite this limitation, it is likely that the 2000 Census did capture many informal workers. Throughout this report the region of study is limited to the Los Angeles-Long

* Informal jobholders are defined herein as workers who self-reported in the 2000 Census that they were privately employed during the enumeration period, but who have no official earnings in either Q1 or Q2 of 2000 according to LEHD data. Formal workers were those classifying themselves as employed for which at least one quarter of earnings was reported through LEHD data in the first half of 2000.

Beach Primary Metropolitan Statistical Area (PMSA).*

In this analysis of informal jobholders there are three key questions that should be resolved: (1) Are informal workers adequately captured by LEHD data? (2) Is the informal jobholder population sufficiently large that it cannot be ignored in labor-workforce analyses and transportation modeling? (3) Are the personal, employment, commuting, and spatial characteristics of this population sufficiently different from formal jobholders that they need to be explicitly factored into labor-workforce analyses and transportation modeling? The first question is immediately answered by the fact that 13.8% of the privately employed workers living in Los Angeles County were classified as informal jobholders because they did not appear as earning wages in the first two quarters of LEHD UI data.[†] This number, which is in the range of previously reported data reviewed in the next section, means that the LEHD data, by itself, is not adequately capturing informal workers. Again, the size of the informal jobholder pool suggests that this group of workers is of significant importance for future analyses of the metropolitan labor market and modeling of the commute-to-work patterns. Finally the overall commute and spatial patterns among formal and informal jobholders are somewhat different, particularly with respect to commuting and spatial patterns. Together these data suggest that informal jobholders should be explicitly included in future transportation analyses and modeling.

BACKGROUND

For the purposes of this technical report, informal workers are those who work “off-the-book.” That is, their employment and earnings are not recorded in official administrative records, which is the reason why they do not appear in the LEHD UI data. Another term for this segment of the economy is the “underground economy” or “informal economy,” which the state of California defines as the economic sector that evades or falls outside governmental oversight and regulation. The “underground economy” includes both illegal activities such as drug trafficking and money laundering, and semi-legal activities such as employment in the garment and

* The Los Angeles-Long Beach Primary Metropolitan Statistical Area (PMSA) and Los Angeles County are coterminous and lie completely within the five-county Los Angeles-Riverside-Orange County Consolidated Metropolitan Statistical Area (CMSA).

[†] The estimates in this report are based on responses from a sample of the population. As with all surveys, estimates may vary from the actual values because of sampling variation or other factors. All statements made in this report have undergone statistical testing and are significant at the 90-percent confidence level, unless otherwise noted.

construction industries where workers are paid “under the table” (Chiang, 1998). While criminal activities garner much attention, the majority of goods and services produced by the “underground economy” are legal (Barber, 2003).

What is known about informal workers comes from studies of underground workers. These workers earn only about half of what a worker earns in the formal economy, but “off-the-book” employment could also involve better paid jobs (Marcelli, 2001; Ong and Rickles, 2004). For example, a construction company might pay its employees \$20 per hour under-the-table to avoid high worker’s compensation insurance. This population makes up a disproportionate share of the informal economy. Furthermore, undocumented workers (who are likely to be recent immigrants and have limited English proficiency) may tend to gravitate to the underground and low-wage economies for work. The U.S. Citizenship and Immigration Services (née INS) estimates that there were about 7.0 million undocumented aliens residing in the United States in 2000 (INS, 2003).

By their very nature, it is difficult to estimate the size of the underground economy and the informal labor force. Estimates of the relative size of the informal or underground sector in western economies range from 3% to 40% of GDP, and perhaps 10% of the U.S. GDP (Barber, 2003). The available data indicate that the underground economy in this country more than doubled from 1970 to 2000, from 4% of GDP in 1970 to 9% in 2000. Given the estimates of the size of the informal or underground economy, it is likely that informal workers comprise at least a tenth of the nation’s labor force. In states like California, which has a disproportionate number of immigrants and industries that pay workers in cash, the relative size of the informal workforce is likely to be substantially higher. According to one estimate, informal workers represented approximately 17% of California’s workforce in 1997-1999 (Marcelli, 2001). In Los Angeles County, the percentage may be even higher because about a quarter of all workers are paid in cash (Campbell, 2003). While informal workers are concentrated in several low-wage sectors (e.g., garment, food services), they are also present in other sectors such as high-tech, construction, and entertainment.

None of the literature reviewed has analyzed the relationship between the spatial distribution of

place of residence and employment of informal jobholders or of the commuting patterns of those workers. However, given the reported differences in jobholder status across socioeconomic groups, it is expected that some spatial differences would also exist.

DATA, SAMPLE, AND METHODS

The Census Bureau's Longitudinal Employer-Household Dynamics (LEHD) Program works with partner states to gather quarterly Unemployment Insurance (UI) data and develop data products for the states' use based upon that data. These UI data are collected by the states from employers and the datasets thus provide data primarily on jobs rather than on individuals. However, each worker is tagged with a Protected Identification Key (PIK) representing one individual* and thus can be linked to other Census datasets following appropriate confidentiality processing so that socioeconomic factors can be matched with the quarterly UI earnings records.

The socioeconomic Census data used in this analysis comes from the 2000 Decennial Census Long Form (LF) dataset, which is an average 1-in-6 sample of the entire U.S. population. Each record represents one respondent from the Decennial Census and contains socioeconomic data as well as commuting data and geographical data on the individual's residence and place of work. It also includes a weighting factor that allows the full population to be imputed.

By using LEHD data matched to Census LF data, we were able to identify informal jobholders and develop some insights into the characteristics of their jobs by industry and earnings, as well as places of work and residence. Since in the strictest sense informal workers should not appear in LEHD data, we used the LF data and the lack of matching records in LEHD to identify these informal workers. Also, because we included those workers who had earnings in only one of the first two quarters in the "formal" category and because we performed no comparison between LEHD earnings and reported 1999 earnings from the LF, this method tends to generate a conservative estimate of the informal jobholder population. Used as a "control" was the formal population, which identified having earnings in at least one of the first two quarters as reported through the UI system. For both formal and informal jobholders the universe within the LF

* The PIK/SSN substitution process is the same for all Census data and is handled in-house by a single unit at the Census Bureau to ensure protection of confidential data.

sample population was the same: individuals who identified themselves as currently employed by private firms and at-work during the enumeration period. We purposefully exclude public-sector employees and the self-employed.

Four analyses were performed with the information generated by the above dataset, and the Los Angeles metropolitan area (see footnote on page 5) was used as a case study for the four analyses. The first analysis examines and compares the social characteristics of informal jobholders and formal jobholders by using the merged data from the LF sample of the 2000 Census. By merging individual-level data from the 2000 Census, it is possible to examine the individual and household characteristics associated with being an informal jobholder. Specifically we considered sex, age, educational attainment, race/ethnicity, and nativity.

The second analysis focuses on employment outcomes, particularly sectoral (industrial) distribution and earnings. Total annual earnings from 1999, rather than quarterly earnings from LEHD, were used because they provide information on both those workers identified as formal as well as informal. Because this earnings data is from 1999 rather than 2000, it should be seen as the “earnings potential” for each worker in 2000.

The third analysis examines and compares the travel patterns of informal jobholders and formal jobholders by using the LF data on the commute to work, specifically travel mode and travel time. A comparison of distance between residence and work (or origin-destination) uses the LF’s locational information. Because of uncertainty in the geocoding results of the locations – particularly the POWs – jobholders’ residences and the firm locations were aggregated at the tract level and all distances were calculated between the tract centroids.

The fourth analysis examines and compares the spatial distribution of POWs and PORs for both informal and formal jobholders. Making use of the origin-destination matrix (by tract), which was constructed and summarized in the previous analysis, we compared the spatial distribution of PORs against POWs for each type of jobholder. Two analytical methods were used to compare these spatial distributions. The first was a simple correlation of tract origins and destinations. The second was the calculation of dissimilarity indices, a method which is

described below.

We calculated dissimilarity indices, which is a widely used measurement in the study of residential segregation (Iceland et al., 2002), as one method of comparing the spatial distribution of formal and informal jobholders. The index ranges from 0 to 100 and roughly indicates the percentage of a group that would have to move to achieve full integration across the universe of geographical units with another group. The following is an example of its application to the analysis in this technical report.

$$DI = 100 \cdot \sum_i \frac{1}{2} \left| \frac{p_i}{P} - \frac{q_i}{Q} \right|$$

where DI = Dissimilarity Index (0-100)

p_i = Count of first population in geographical unit i .

P = Count of first population across the universe of geographical units.

q_i = Count of second population in geographical unit i .

Q = Count of second population across the universe of geographical units.

In this report several different dissimilarity indices were calculated (e.g., between the POR of informal jobholders and the POR of formal jobholders). The geographical unit of analysis for these calculations is the Census Tract, and the universe of units is Los Angeles County.

FINDINGS

Overall, 13.8% of our sample were identified as informal jobholders. This is within the reported ranges of the literature and is comparable to expectations of the specific workforce composition of Los Angeles County. Of course, the sample does not include those workers in informal jobs who either were not enumerated by the Census or who were reluctant to report their earnings to the Census. Nor does it attempt to determine the size of the population that used legal Social Security Numbers (SSNs) in an illegal manner (e.g., workers without SSNs using someone else's SSN for required IRS forms). Throughout the subsequent section, informal workers may also be identified as those who worked 0 quarters in reference to their absence from quarterly UI administrative data. The final sample comprising merged and weighted records from LEHD and

the LF includes 2,397,086 observations. The demographic composition of this sample is summarized in Table 1.

Table 1: Demographic Characteristics			
	Total Sample	Formal JHs	Informal JHs
	[%]	[%]	[%]
Sex			
Male	54.0	54.0	54.1
Female	46.0	46.0	45.9
Age			
16-24	15.3	15.4	14.6
25-44	54.4	55.3	49.0
45-64	28.0	27.4	31.3
65+	2.4	1.9	5.1
Education			
Less than HS	20.9	20.0	26.9
HS	18.7	18.8	18.1
Some College	32.9	33.7	27.6
BA/BS+	27.5	27.5	27.4
Nativity			
U.S. Born	60.4	61.6	53.5
Naturalized	18.8	19.0	18.1
Non-citizen	20.7	19.5	28.4
Race/Ethnicity			
NH White	39.3	39.8	36.6
Black	8.0	7.8	9.1
Asian	13.4	13.4	13.5
Hispanic	33.7	33.6	34.8
Others	5.5	5.4	6.1

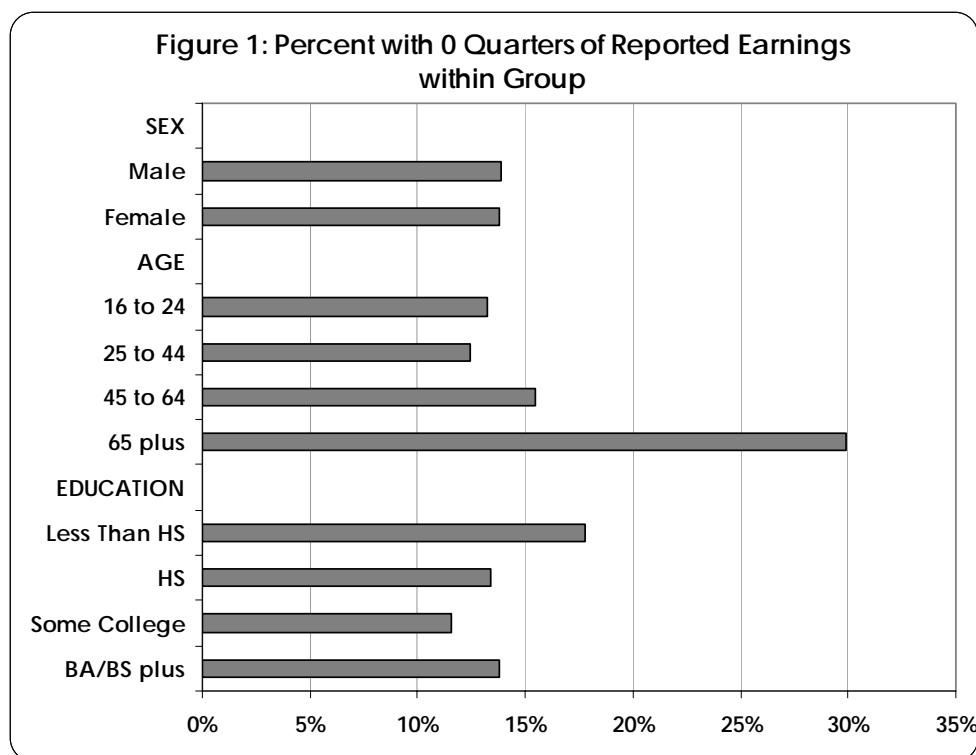
Sources: U.S. Census Bureau, Census 2000; U.S. Census Bureau, LEHD Program 2005.

Notes:

1. Total % for each group may not equal 100% due to rounding.
2. Certain values in this table may not be significantly different from one another.
3. Data based on sample. For information on confidentiality protection, sampling error, and definitions, see <<http://www.census.gov/prod/cen2000/doc/sf3.pdf>>.
4. For further information on confidentiality protection and definitions with respect to UI data, see <<http://lehd.did.census.gov/led/library/techpapers/tp-2006-01.pdf>>.

Consistent with the literature, a larger percentage of the informal jobholders are men, and they are primarily aged 25-64. The male/female ratio of informal workers is consistent with their shares of the total workforce population. While overall percentages of informal jobs are relatively consistent across the educational attainment groups (approximately 27% for those with less than a high school degree, as well as those with some college or more advanced degrees), some differences can be detected by comparing to educational attainment within the total population. Whereas those with less than a high school education are overrepresented as informal

jobholders (26.9% informal to 20.9% of the total workforce), those who have some college education are underrepresented within the informal jobholder group (27.6% compared with 32.9% of all JHs).



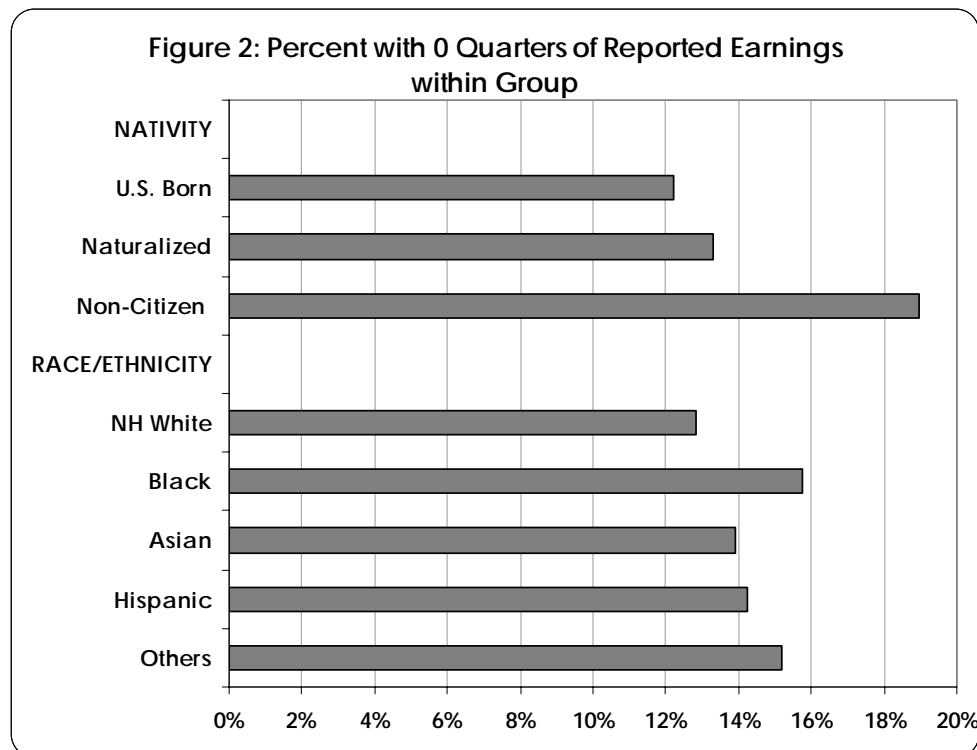
Sources: U.S. Census Bureau, Census 2000; U.S. Census Bureau, LEHD Program 2005.

Notes:

1. Certain values in this figure may not be significantly different from one another.
2. Data based on sample. For information on confidentiality protection, sampling error, and definitions, see <http://www.census.gov/prod/cen2000/doc/sf3.pdf>.
3. For further information on confidentiality protection and definitions with respect to UI data, see <http://lehd.did.census.gov/led/library/techpapers/tp-2006-01.pdf>.

There also appears to be a bias toward older workers when we consider each group's share of informal workers relative to its share within the full sample population. The age group of those between 25-44 are underrepresented as informal workers with 49.0% of informal jobholders in this age group relative to 54.4% of the total workforce. Also underrepresented is the youngest age group (16-24) with 14.6% of the informal jobholders compared to 15.3% of all workers. These differences are offset by overrepresentation within the other two age groups: ages 45-64 with 31.3% informal vs. 28.0% total and ages 65+ with 5.1% informal vs. 2.4% total.

These patterns are also apparent when we compare the percentages of each demographic class that qualify as informal. (See Figure 1.) While males and females have about the same share of their populations working in the informal sector, there is a clear bias towards older workers (aged 65+), about 30% of whom are working informally. The distribution of shares by educational attainment suggests what was noted in the literature review: there is a better chance of being employed informally with less education, but those with college degrees have a substantial chance (approximately 14%) of being in the informal sector.



Sources: U.S. Census Bureau, Census 2000; U.S. Census Bureau, LEHD Program 2005.

Notes:

1. Certain values in this figure may not be significantly different from one another.
2. Data based on sample. For information on confidentiality protection, sampling error, and definitions, see <http://www.census.gov/prod/cen2000/doc/sf3.pdf>.
3. For further information on confidentiality protection and definitions with respect to UI data, see <http://lehd.did.census.gov/led/library/techpapers/tp-2006-01.pdf>.

While U.S.-born workers held the largest share of informal jobs (53.5% of all informal workers), their share is smaller compared with their overall population within the full workforce. That is, U.S.-born jobholders make up 60.4% of the Los Angeles County workforce, but only account for 53.5% of its informal jobholders. In turn, and as might be expected, this difference is made up mostly within the non-citizen population, which represents 20.7% of the local population but

28.4% of the informal jobholder population. Meanwhile, naturalized citizens have a slightly smaller proportion as part of the informal population (18.1%) than of the overall workforce (18.8%). These differences among groups are compared directly in Figure 2, which shows the relative share of each group classified as informal workers, and clearly non-citizens are substantially more likely to be informal jobholders.

By race, this effect of disproportionate shares of the informal population is not as striking. For example, non-Hispanic whites make up 39.3% of the privately employed workforce but only 36.6% of the informal jobholder population. Hispanics, who by popular conception might be expected to have a significantly larger share of the informal population, actually show only a slightly higher share (33.7% of the total workforce vs. 34.8% of informal workers). Again, the direct comparisons are shown in Figure 2.

Table 2: Job Composition by Industry		
	Informal JHs [%]	Formal JHs [%]
Mgmt., Bus., Fin. Svcs.	6.3	4.7
Comp., Arch., Eng., Sci.	4.3	5.1
Comm., Ed., Law, Arts, Sports, Media	1.9	3.9
Health, Protective Svcs.	4.8	9.7
Food, Sales, Maint., Pers. Svcs.	7.6	11.8
Office, Admin.	5.2	6.8
Farm, Constr., Mining	12.2	18.1
Install., Maint., Repair, Prod.	28.6	18.8
Add'l Prod.	19.8	18.7
Trans., Mil., Unempd.	9.3	2.5
Total	100.0	100.0

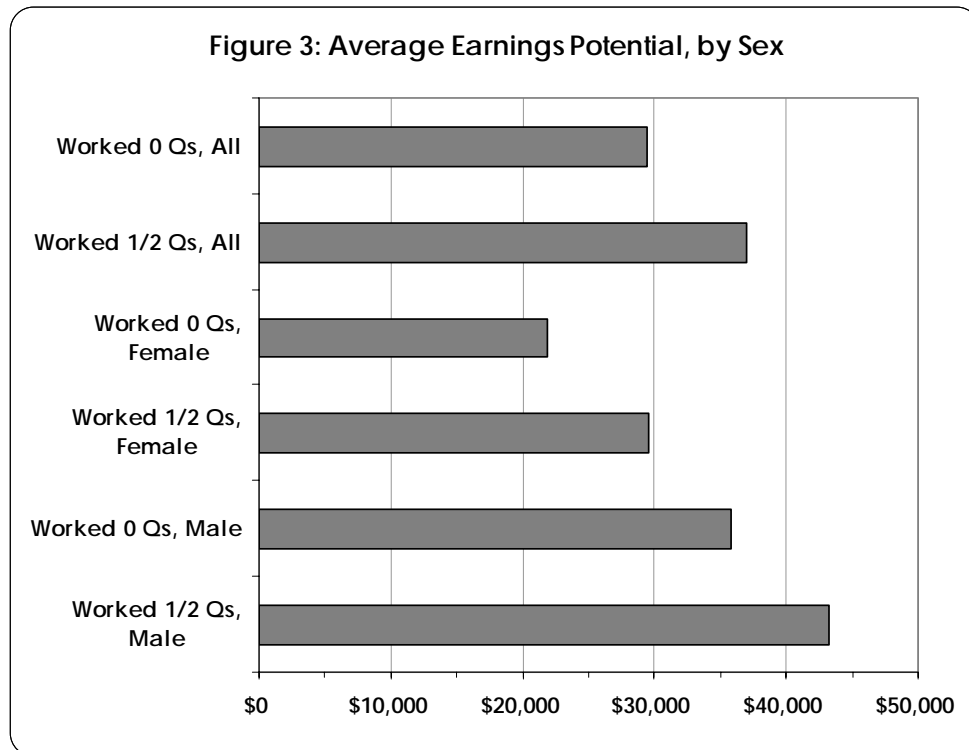
Sources: U.S. Census Bureau, Census 2000; U.S. Census Bureau, LEHD Program 2005.

Notes:

1. Total % for each group may not equal 100% due to rounding.
2. Certain values in this table may not be significantly different from one another.
3. Data based on sample. For information on confidentiality protection, sampling error, and definitions, see <<http://www.census.gov/prod/cen2000/doc/sf3.pdf>>.
4. For further information on confidentiality protection and definitions with respect to UI data, see <<http://lehd.did.census.gov/led/library/techpapers/tp-2006-01.pdf>>.

Table 2 compares the distribution of jobs by industrial sector using the Census 2000 Industry Codes by 1-digit groupings. Relative to formal jobholders, informal jobholders are more likely to be in the installation, maintenance, and production sector as well as the transportation sector than in health and protective services sector or the farm/construction/mining sector. It is interesting to note the larger share of informal workers in the management, business, and financial services

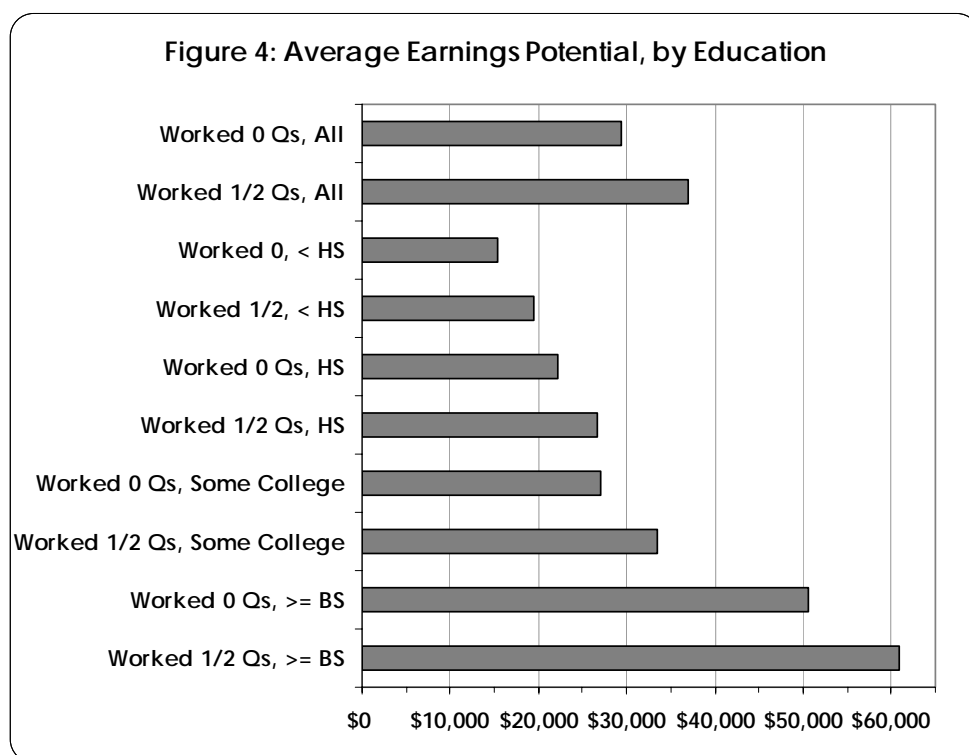
sector. This may be due to misreporting by the self-employed who did not identify themselves as such; however, this problem is beyond the scope of this report.



Sources: U.S. Census Bureau, Census 2000; U.S. Census Bureau, LEHD Program 2005.

Notes:

1. Certain values in this figure may not be significantly different from one another.
2. Data based on sample. For information on confidentiality protection, sampling error, and definitions, see <http://www.census.gov/prod/cen2000/doc/sf3.pdf>.
3. For further information on confidentiality protection and definitions with respect to UI data, see <http://lehd.did.census.gov/led/library/techpapers/tp-2006-01.pdf>.



Sources: U.S. Census Bureau, Census 2000; U.S. Census Bureau, LEHD Program 2005.

Notes:

1. Certain values in this figure may not be significantly different from one another.
2. Data based on sample. For information on confidentiality protection, sampling error, and definitions, see <<http://www.census.gov/prod/cen2000/doc/sf3.pdf>>.
3. For further information on confidentiality protection and definitions with respect to UI data, see <<http://lehd.did.census.gov/led/library/techpapers/tp-2006-01.pdf>>.

Because earnings were, by definition, not available from UI data for informal workers, we used total reported 1999 wages from the LF data as a proxy for 2000 “earnings potential” in order to keep the earnings comparisons consistent between the formal and informal groups. Even with this proxy variable, it is clear from the data (see Figures 3 and 4) that informal jobholders earn less – in some cases as much as 20% less – than their peers in terms of both sex as well as educational attainment, though this difference is less than the 2:1 pay ratio suggested in the reviewed literature. It is interesting to note, though, that on average men working at informal jobs earn more than women working at formal jobs, a fact that may reflect on the different sectoral/pay-rate composition of jobs in each class that are open to men and women. Percentage-wise by educational attainment, the earnings ratio is largest for those with the least education (21.1% lower earnings for informal workers with less than a high school education), and as a result the least earnings. The rate is also high for workers with some college education who earn

19.5% less as informal jobholders than in the formal sector. For the other two educational groups the earnings penalties for being an informal jobholder are closer to 17%.

Table 3: Commute Characteristics		
	Informal JHs [%]	Formal JHs [%]
Commute Mode		
Drive Alone	64.1	76.5
Carpool	16.2	14.0
Public Transit	7.3	4.9
Other	12.4	4.6
Commute Time		
0 to 14	28.9	20.3
15 to 29	31.6	34.7
30 to 59	29.5	33.9
60 plus	10.0	11.1
Commute Distance		
Same Tract	13.0	3.7
5 miles or less	32.9	33.3
5 to 15 miles	31.4	40.5
Over 15 miles	22.6	22.5

Sources: U.S. Census Bureau, Census 2000; U.S. Census Bureau, LEHD Program 2005.

Notes:

1. Certain values in this table may not be significantly different from one another.
2. Data based on sample. For information on confidentiality protection, sampling error, and definitions, see <<http://www.census.gov/prod/cen2000/doc/sf3.pdf>>.
3. For further information on confidentiality protection and definitions with respect to UI data, see <<http://lehd.did.census.gov/led/library/techpapers/tp-2006-01.pdf>>.

Formal workers enjoy a distinct advantage over informal workers in the use of a car for commuting purposes with 76.5% of formal workers driving alone compared to only 64.1% of informal workers. (See Table 3.) Consequently, informal workers are more likely to use other modes of transit including carpools and public transportation. One obvious discrepancy is the “Other” class commute mode which includes 12.4% of informal jobholders but only 4.6% of formal jobholders. This is likely linked to the fact that many informal workers have no commute (or live and work in the same tract), as is discussed below.

At the same time informal workers appear to spend less time commuting as well as having fewer long commute trips (specifically when comparing Same Tract commutes to those of 5-15 miles). These numbers are somewhat skewed by the significant difference in the share of informal jobholders working at home compared to formal jobholders. 5.1% of informal jobholders report their commute time as 0 minutes whereas only 1.1% of formal jobholders report that case.

However, even after removing those informal jobholders with no commutes from the tabulations, the remaining informal workers still have shorter commutes than their formal sector counterparts. As a result a similar skewing effect would also appear in the distance-to-work data, but the overall patterns hold.

Table 4: Dissimilarity Indices Among Informal/Formal JH Classes

Group Comparison	Dissimilarity Index
Informal JH POW vs. Formal JH POW	29.9
Informal JH POR vs. Formal JH POR	15.7
Informal JH POW vs. Informal JH POR	37.7
Formal JH POW vs. Formal JH POR	52.9

Sources: U.S. Census Bureau, Census 2000; U.S. Census Bureau, LEHD Program 2005.

Notes:

1. Certain values in this table may not be significantly different from one another.
2. Data based on sample. For information on confidentiality protection, sampling error, and definitions, see <<http://www.census.gov/prod/cen2000/doc/sf3.pdf>>.
3. For further information on confidentiality protection and definitions with respect to UI data, see <<http://lehd.did.census.gov/led/library/techpapers/tp-2006-01.pdf>>.

Table 4 lists the various dissimilarity indices among places of residence (PORs) and places of work (POWs) for both the formal and informal jobholder classes within this study. It is clear by the low dissimilarity index (15.7) that informal and formal jobholders tend to be well-integrated in their places of residence. Their places of work are also somewhat integrated with a DI of 29.9. The least integrated groupings are between formal jobholders' PORs and their POWs (DI of 52.9). Compared with this, informal jobholders are better integrated, which results in part from the number of informal jobholders identified above with no commutes.

Table 5: Correlation Coefficients among Informal/Formal JH Classes

	Formal JH POR	Informal JH POW	Formal JH POW
Informal JH POR	0.609 (<0.0001)	0.270 (<0.0001)	0.109 (<0.0001)
Formal JH POR		0.075 (0.0011)	0.127 (<0.0001)
Informal JH POW			0.657 (<0.0001)

Sources: U.S. Census Bureau, Census 2000; U.S. Census Bureau, LEHD Program 2005.

Notes:

1. Certain values in this table may not be significantly different from one another.
2. Data based on sample. For information on confidentiality protection, sampling error, and definitions, see <<http://www.census.gov/prod/cen2000/doc/sf3.pdf>>.
3. For further information on confidentiality protection and definitions with respect to UI data, see <<http://lehd.did.census.gov/led/library/techpapers/tp-2006-01.pdf>>.
4. Standard errors were calculated using SAS 9 and adjusted using the Census 2000 long form design factors.

As can be seen in Table 5, the correlation data for PORs and POWs aggregated at the tract level support the DI calculations with high correlation coefficients between PORs of formal and informal workers as well as between POWs for the two classes of workers. The statistical significance levels are in parentheses. In turn the coefficients between PORs and POWs are low, which is unsurprising given the dissimilarity indices above. However, the coefficients are not negative, which suggests that jobs and homes are not as polarized as one might expect in another city with a standard polar/CBD structure. One other note of interest is that while the correlation coefficient of POWs between informal jobholders and formal jobholders is higher than that for PORs, the relationship is reversed when we consider the DIs – that is, PORs are better distributed, which is not what one would expect given the correlation coefficients.

CONCLUSION

After considering the results of this analysis, we must return to the three primary questions posited in the first section of this report: (1) Are informal jobholders adequately captured in the LEHD data? (2) Is the pool of informal jobholders too large to be ignored in further analyses? (3) Is the pool of informal jobholders too different to be ignored in further analyses? Immediately, the first question is answered through the very nature of this analysis. By definition, informal workers are not being captured by the LEHD data. However, without the LEHD UI data, it would be impossible to make the classifications of formal and informal because the appropriate data is not available in the 2000 Long Form collection.

In addressing the second question, inspection of the relative size of the informal jobholder pool (13.8%) suggests this group is definitely of the size that would warrant the additional work and effort required to incorporate them explicitly into future analyses of the metropolitan labor market and into models of the commute-to-work patterns. As has been mentioned, our class of informal jobholders is a conservative estimate because of reporting and coverage issues, and so the share and importance of this group might well be greater than stated in this data. Even so, the percentage calculated within this report lies within reported ranges for informal workers and is comparable with estimates of informal workers in the Los Angeles area.

In terms of social characteristics, formal and informal classes of workers have some moderate differences. In terms of sex, the two classes are essentially equivalent, but for age and educational attainment the tendency is for older, less educated workers to hold informal jobs – a set of relations that may signify a split in the types of informal jobs being held. In terms of nativity, informal jobs tend to go to non-citizens over U.S.-born workers, though the relationships do not hew particularly closely to expectations in terms of race/ethnicity. The differences in industrial breakdown among the classes are also as one might expect, and when the job classes are compared by earnings there are some clear and obvious differences, with formal workers often having significantly higher earnings potential.

When we compare the two groups by commuting patterns, we find that some differences remain, especially considering the rate of driving alone to work as well as the distance/time to work. Some of this can be explained by the relatively high rate of informal workers who have no commute, though it is not all of the difference between the two groups. Finally, the patterns in spatial distribution are somewhat less distinct. Formal and informal jobholders tend to be more integrated in where they live than where they work, and there is some suggestion in the data that the places of work and residence are less likely to be integrated for formal workers than for informal workers. Overall, there are important differences in socioeconomic, commuting and spatial patterns, and informal workers should definitely be considered as a separate group for modeling and analysis purposes, which was the point of the third thematic question.

One significant caveat that must be mentioned is the likelihood that these results are strongly linked to the urban structure and economic spatial distribution of Los Angeles itself. As a multipolar city-region with complex commute patterns, these results may not be as applicable for those cities with a more classical polar/CBD (Central Business District) structure. Another factor germane to this report and worth further study in comparisons with other metropolitan areas is the effect of recent immigration and low-wage workers, which should be examined in a greater depth but which is beyond the scope of this project.

As for the analysis itself that was used for this study, the methodology of linking UI data and socioeconomic data from the Census LF appears to offer a rich resource in studying transportation among particular classes of workers. In addition, it appears that this methodology can be easily ported to analyses involving the American Community Survey, which has a smaller sample size but more current statistics than the Decennial Census.

ACCURACY OF THE ESTIMATES

The data contained in this report are based on the sample of households who responded to the Census 2000 long form. Nationally, approximately one out of every six housing units was included in this sample. As a result, the sample estimates may differ somewhat from the 100-percent figures that would have been obtained if all housing units, people within those housing units, and people living in group quarters had been enumerated using the same questionnaires, instructions, enumerators, and so forth. The sample estimates also differ from the values that would have been obtained from different samples of housing units, and hence of people living in those housing units, and people living in group quarters. The deviation of a sample estimate from the average of all possible samples is called the sampling error.

In addition to the variability that arises from the sampling procedures, both sample data and 100-percent data are subject to nonsampling error. Nonsampling error may be introduced during any of the various complex operations used to collect and process data. Such errors may include: not enumerating every household or every person in the population, failing to obtain all required information from the respondents, obtaining incorrect or inconsistent information, and recording information incorrectly. In addition, errors can occur during the field review of the enumerators' work, during clerical handling of the census questionnaires, or during the electronic processing of the questionnaires.

While it is impossible to completely eliminate error from an operation as large and complex as the decennial census, the Census Bureau attempts to control the sources of such error during the data collection and processing operations. The primary sources of error and the programs instituted to control error in Census 2000 are described in detail in Summary File 3 Technical Documentation under Chapter 8, "Accuracy of the Data," located at <http://www.census.gov/prod/cen2000/doc/sf3.pdf>.

Nonsampling error may affect the data in two ways: (1) errors that are introduced randomly will increase the variability of the data and, therefore, should be reflected in the standard errors; and (2) errors that tend to be consistent in one direction will bias both sample and 100-percent data in that direction. For example, if respondents consistently tend to underreport their incomes, then

the resulting estimates of households or families by income category will tend to be understated for the higher income categories and overstated for the lower income categories. Such biases are not reflected in the standard errors.

All statements in this Working Paper have undergone statistical testing and all comparisons are significant at the 90-percent confidence level, unless otherwise noted. The estimates in tables, maps, and other figures may vary from actual values due to sampling and nonsampling errors. As a result, estimates in one category used to summarize statistics in the maps and figures may not be significantly different from estimates assigned to a different category. Standard errors for the Dissimilarity Index estimates were calculated using derived methods based upon folded normal distribution theory. See Elandt, 1961. Further information on the accuracy of the data is located at <<http://www.census.gov/prod/cen2000/doc/sf3.pdf>>. For further information on the computation and use of standard errors, contact the Decennial Statistical Studies Division at 301-763-4242.

REFERENCES

- Abowd, John M., Bryce E. Stephens, Lars Vilhuber, Fredrik Andersson, Kevin L. McKinney, Marc Roemer, and Simon Woodcock. 2005. "The LEHD Infrastructure Files and the Creation of the Quarterly Workforce Indicators." *LEHD Technical Paper*. No. TP-2006-01. Washington, D.C.: U.S. Census Bureau.
- Barber, D.A. 2003. "The 'New' Economy? Fast-growing 'Underground Economy' Plays a Larger Role in Tough Times." *Tucson Weekly*, 2 January. Accessed 14 November 2003 <<http://www.tucsonweekly.com/gbase/currents/Content?oid=oid:46422>>.
- Campbell, D. 2003. "With Pot and Porn Outstripping Corn, America's Black Economy is Flying High: Illegal Migrants Provide the Muscle for US Black Market." *The Guardian*, 2 May. Accessed 14 November 2003 <<http://www.guardian.co.uk/usa/story/0,12271,947880,00.html>>.
- Chiang, J. 1998. *Underground Economy Constituent Report* 1(1). Accessed 14 November 2003 <<http://www.boe.ca.gov/members/chiang/constrpt/ueconstrp.htm>>.
- Economic Development Department. "Employment Tax: Underground Economy Operations." *Employment Development Department*. Accessed 14 November 2003 <<http://www.edd.ca.gov/taxrep/txueoind.htm>>.
- Elandt, R.C. "The Folded Normal Distribution: Two Methods of Estimating Parameters from Moments." *Technometrics*. 3(4): 551-562.
- Iceland, John, Daniel H. Weinberg, and Erika Steinmetz. 2002. "Racial and Ethnic Residential Segregation in the United States: 1980-2000." *U.S. Census Bureau Series CENSR-3*. Washington, D.C.: U.S. Government Printing Office.
- Iwata, E. 2003. "Financial Woes Mire California's Once Strong Economy." *USA Today*, 12 April.
- Levine, D. S. 2003. "Underground 'Squeezes' Honest Firms: Employer Fraud Pressures Businesses Trying to Stay Legit." *San Francisco Business Times*, 25 August. Accessed 14 November 2003 <<http://sanfrancisco.bizjournals.com/sanfrancisco/stories/2003/08/25/focus4.html>>.
- Marcelli, E. "Unauthorized Mexican Immigration, Day Labor, and Other Lower-wage Informal Employment in California." *Regional Studies*.
- Ong, Paul, Jordan Rickles, Amy Ford, Matthew Graham, Todd Nelson, Paul Castro, and Hyun-Gun Sung. 2003. *Analysis of the California Labor and Workforce Development Agency's Enforcement of Wage and Hour Laws: Final Report*. Los Angeles: UCLA Ralph & Goldy Lewis Center for Regional Policy Studies.

- U.S. Department of Justice. 2003. "INS Releases Updated Estimates of U.S. Undocumented Resident Population: 2000 Census Data and New Methodology Improve Estimates. *U.S. Citizenship and Immigration Services*. Revised 21 March 2005. Accessed 19 October 2005 <<http://uscis.gov/graphics/publicaffairs/summaries/undocres.htm>>.
- Wu, Jeremy, Paul Ong, and Marc Roemer. "Supplementing the American Community Survey: The LEHD Program." Presented at the Conference on Census Data for Transportation Planning organized by the Transportation Research Board, held at the National Academies Beckman Center, Irvine, CA, 12 May 2005.